(2)

A CENTRAL LIMIT THEOREM FOR AUTOREGRESSIVE INTEGRATED MOVING AVERAGE PROCESSES

J. E. Angus

Report No. 92-15

NAVAL HEALTH RESEARCH CENTER
P.O. BOX 85122
SAN DIEGO, CALIFORNIA 92186-5122

NAVAL MEDICAL RESEARCH AND DEVELOPMENT COMMAND
BETHESDA, MARYLAND

A CENTRAL LIMIT THEOREM FOR AUTOREGRESSIVE INTEGRATED MOVING AVERAGE PROCESSES

John E. Angus

Department of Mathematics
The Claremont Graduate School
143 E. Tenth Street
Claremont, CA 91711-3988

and

Medical Information Systems and Operations Research Department
Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

August 1992

Report No. 92-15, supported by the Naval Health Research Center under the American Society for Engineering Education Navy Summer Faculty Research Program. The views expressed in this article are those of the author and do not reflect the official policy or position of the Department of Defense, nor the U.S. Government. Approved for public release, distribution unlimited.

SUMMARY

A central limit theorem for normalized sums of random variables that form an autoregressive integrated moving average (ARIMA) process is developed. The need for such a limit theorem is discussed in connection with modeling total compensation costs associated with insurance or medical claims.

A CENTRAL LIMIT THEOREM FOR AUTOREGRESSIVE INTEGRATED MOVING AVERAGE PROCESSES

JOHN E. ANGUS

Department of Mathematics
The Claremont Graduate School
143 E. Tenth Street
Claremont, CA 91711

1. Introduction

In order to make the discussion reasonably self contained, it is necessary to introduce autoregressive integrated moving process and related concepts. For a complete development and discussion of the mathematics and applications of these processes, refer to Brockwell and Davis (1992), upon which the present notation and discussion is based.

Consider a fixed probability space $(\Omega, \mathfrak{F}, P)$ on which all subsequent random variables will be defined. A collection of random variables $\{Z_t, t=0, \pm1, \pm2, ...\}$ is said to be a white noise process if $EZ_t=0$, $E(Z_t^2)=\sigma^2$, for all t, and $E(Z_tZ_s)=0$ for all s, t with $s\neq t$. This is denoted by $\{Z_t\}\sim WN(0,\sigma^2)$. If the Z_ts are also independent and identically distributed, this is indicated by $\{Z_t\}\sim IID(0,\sigma^2)$. $\{Y_t, t=0, \pm1, \pm2, ...\}$ is said to be an autoregressive moving average process with autoregressive order p and moving average order q, denoted by $\{Y_t\}\sim ARMA(p,q)$, if Y_t satisfies a set of difference equations of the form (the φ_is and θ_is are fixed real constants)

$$Y_t-\varphi_1Y_{t-1}-\varphi_2Y_{t-2}-...-\varphi_pY_{t-p} = Z_t+\theta_1Z_{t-1}+\theta_2Z_{t-2}+...+\theta_qZ_{t-q} \quad (1)$$

for all integer t, where $\{Z_t\}\sim WN(0,\sigma^2)$, and the polynomial $\varphi(z)=1-\varphi_1z-\varphi_2z^2-...-\varphi_pz^p$ has no roots on the unit circle $\{z:|z|=1\}$ in the complex plane. Introducing the back-shift operator B, where $BY_t=Y_{t-1}$, $B^jY_t=Y_{t-j}$, for integer j, (1) can be written compactly as

$$\varphi(B)Y_t=\theta(B)Z_t \quad (2)$$

where $\varphi(z)=1-\varphi_1z-...-\varphi_pz^p$, and $\theta(z)=1+\theta_1z+\theta_2z^2+...+\theta_qz^q$. By definition, $EY_t=0$ for all t. The process $\{Y_t, t=0, \pm 1, \ldots\}$ is said to be an ARMA(p,q) with mean μ if $\{Y_t-\mu, t=0, \pm 1, \ldots\}$ is an ARMA(p,q) process. The condition

$\varphi(z)\neq 0$ for $|z|=1$

insures that the process $\{Y_t, t=0, \pm 1, \ldots\}$ satisfying (1) is stationary, which means that the autocovariance function $\gamma(s,t)=\mathrm{cov}(Y_t, Y_s)$ depends only on $|t-s|$, so that it can be expressed as

$$\gamma_Y(h) \equiv \mathrm{cov}(Y_t, Y_{t+h}) \tag{3}$$

without ambiguity. Moreover, if $\varphi(z)\neq 0$ for $|z|=1$, then the difference equations (1) have a unique solution given by

$$Y_t=\sum_{j=-\infty}^{\infty}\psi_j Z_{t-j} \tag{4}$$

where the series converges almost surely and in mean square, the coefficients $\{\psi_j\}$ satisfy

$\sum_{j=-\infty}^{\infty}|\psi_j|<\infty$,

and $\{\psi_j\}$ are the coefficients in the Laurent expansion $\theta(z)/\varphi(z) = \sum_{j=-\infty}^{\infty}\psi_j z^j$, valid for z satisfying $r<|z|<1/r$, for some $r\in(0,1)$. In some applications, it is desirable to require that the representation (4) have $\psi_j=0$ for $j<0$, so that Y_t is expressed as a linear combination of current and past Z_ts. This is true if $\varphi(z)\neq 0$ for $|z|\leq 1$, i.e. all the roots of $\varphi(\bullet)$ lie outside the unit circle in the complex plane. Such a process is then called a causal ARMA(p,q). It can be shown that as long as $\varphi(z)\neq 0$ for $|z|=1$, an ARMA(p,q) process always has a causal representation. That is, it is always possible to redefine the white noise process and the polynomial φ so that the process is causal. It will be assumed that all ARMA(p,q) processes discussed herein are causal.

ARMA processes are useful in describing or approximating a wide variety of stationary processes whose autocovariance functions approach zero as the lag approaches infinity. A great many methods have been devised for estimating the orders p and q, and the unknown φ_is and θ_js in (1) for a given set of observations. For example, see Box and Jenkins (1970), Brockwell and Davis (1991), and Priestley (1981). Often however, it is necessary to model various types of nonstationary processes, for example, those with additive components of trend and seasonality. Processes containing polynomial trend and/or periodic behavior can be modeled essentially by allowing the autoregressive polynomial to have one or more roots on the unit circle in the complex plane. One such type of model is called the autoregressive integrated moving average (or ARIMA) process. These are particularly useful for modeling processes without a seasonal component, but which are "explosively" nonstationary, as is the case when the series has a deterministic or stochastic polynomial trend. ARIMA processes are defined as follows.

Define the difference operator $\nabla \equiv (1-B)$, $\nabla^0 \equiv 1$, and $\nabla^j = \nabla(\nabla^{j-1})$ for $j \geq 1$. Let d be a non negative integer. The stochastic process $\{X_t, t=1-d, 2-d, \ldots, 0, 1, 2, \ldots\}$ is called an ARIMA (p,d,q) process if $\nabla^d X_t = Y_t$ where $\{Y_t\}$ is a causal ARMA(p,q) process with mean μ. Thus, for example, if $X_t = A_0 + A_1 t + \ldots + A_{d-1} t^{d-1} + Y_t^*$ where $\{Y_t^*\}$ is a causal ARMA process and the A_is are arbitrary random variables, then $\{X_t\}$ is an ARIMA(p,d,q) for some p and q. This follows easily from the result that $\nabla^d p_t = 0$ for any polynomial $p_t = A_0 + A_1 t + \ldots + A_m t^m$ of degree m<d.

In the next section, the proper centering and normalization of $\sum_{t=1}^{n} X_t$ to achieve an asymptotic normal distribution is studied when $\{X_t\}$ is an ARIMA(p,d,q) process and the white noise process $\{Z_t\}$ appearing in (1) and (2) is actually an $\mathrm{IID}(0,\sigma^2)$ process. Interest in this problem is stimulated by the modeling of medical or insurance claims. A typical model for insurance claims is the so called compound Poisson process. See Prabhu (1980), for example, for extensive discussions of this model in the insurance risk context. Here, claims are generated according to a nonhomogeneous Poisson process $\{N_T, T \geq 0\}$, and successive claim costs are assumed independent of $\{N_T, T \geq 0\}$, and to form a sequence of iid random variables, $\{Y_t, t \geq 1\}$. Thus, total claim costs from the time period [0,T] are given by

$$C_T = \sum_{t=1}^{N_T} Y_t \qquad (5)$$

where a sum with upper index 0 is defined to be 0. If $E(N_T) = m(T)$, $m(T) \uparrow \infty$ as $T \to \infty$, and $\{Y_t\} \sim \mathrm{IID}(\mu, \tau^2)$, then it follows from elementary limit theory that

$$\frac{C_T - m(T)\mu}{\left[m(T)(\mu^2 + \tau^2)\right]^{1/2}} \Rightarrow N(0,1) \qquad (6)$$

as $T \to \infty$, where $N(m, \tau^2)$ denotes a random variable that is normally distributed with mean m and variance τ^2. The notation $X_n \Rightarrow X$ as $n \to \infty$, means $P\{X_n \leq x\} \to P\{X \leq x\}$ as $n \to \infty$, for all x at which the function $x \mapsto P\{X \le x\}$ is continuous. The result (6) allows the distribution of total claim cost C_T to be approximated for large T. This model is important to the insurance industry, since if premiums are collected at a constant rate p per unit time, and the firm starts initially with a cash reserve of c, then the quantity $c + pT - C_T$ represents, in a simplified setting, the monetary reserve of the insurance company at time T, and the first time that this process hits the value zero, the company becomes insolvent.

The model (5) is also plausible for describing medical claims / compensation costs associated with accidents or hazardous materials exposure. In this context, the model (5) can be made more realistic by allowing the claim amounts to be correlated and/or to have a trend. Legal (and other) precedents / interventions and economic factors can affect successive claim costs to the extent that an ARIMA(p,d,q), with suitable p, d, and q, would be a more appropriate model. If this is the case, then in order to develop limit theorems similar to (6), it is necessary to study the asymptotic distribution of $\sum_{t=1}^{n} X_t$, where $\{X_t\}$ is an ARIMA(p,d,q) process. To facilitate this, it will be assumed from now on that the white noise process $\{Z_t\}$ that appears in (1) and (2) is actually an IID sequence, i.e. that $\{Z_t\} \sim \mathrm{IID}(0,\sigma^2)$.

The case where $\{X_t - \mu\} = \{Y_t\} \sim \mathrm{ARMA}(p,q)$ is a special case of Theorem 7.1.2 of Brockwell and Davis (1992), which shows that

$$n^{-1/2} \sum_{t=1}^{n} (X_t - \mu) \rightarrow N(0, v^2) \text{ as } n \rightarrow \infty, \tag{7}$$

where

$$v^2 = \gamma_Y(0) + 2\sum_{h=1}^{\infty} \gamma_Y(h),\ \gamma_Y(h) = \mathrm{cov}(Y_t, Y_{t+h}), \tag{8}$$

provided that $v^2 > 0$. Then the factor $(\mu^2 + \tau^2)$ in (6) becomes $(\mu^2 + v^2)$. Since τ^2 corresponds to $\gamma_Y(0)$, the asymptotic variance of C_T could be larger or smaller than in the IID case, depending on the values of the autocovariances. When $\{X_t\}$ is an ARIMA(p,d,q) process, then $n^{-1/2}$ must be replaced by $n^{-d-1/2}$ as the proper normalization of $\sum_{t=1}^{n} X_t$, as will be seen in the next section.

2. Central Limit Theorem for a Class of ARIMA Processes

Let $\{X_t,\ t=1\text{-}d,\ 2\text{-}d,\ ..,\ 0,\ 1,\ 2,\ ...\}$ be an ARIMA(p,d,q) process, satisfying $\nabla^d X_t = Y_t + \mu$ where $\{Y_t\}$ is a causal ARMA(p,q) process as in (1) and (2) with $\{Z_t\} \sim \mathrm{IID}(0,\sigma^2)$. Define the usual operation of generating factorial polynomials by $k^{(j)} \equiv \prod_{i=1}^{j}(k-i+1) = k(k-1)...(k-j+1)$ for integers j and k with j>0, and $k^{(0)} \equiv 1$. Thus, with k treated as a variable, it follows that for $j \geq 0$,

$$\nabla\,(k+1)^{(j+1)}/[(j+1)!] = k^{(j)}/j!, \tag{9}$$

and

$$\sum_{k=0}^{n} k^{(j)}/j! = \frac{(n+1)^{(j+1)}}{(j+1)!}. \tag{10}$$

The main result is the following theorem, which holds even for the case d=0 with the convention that summations with upper limit 0 are taken to be 0.

Theorem 1. Let $\gamma_Y(\bullet)$ be the autocorrelation function of $Y_t = \nabla^d X_t - \mu$, and suppose that $\gamma_Y(0) + 2\sum_{h=1}^{\infty}\gamma_Y(h) > 0$. Then, as $n \to \infty$, the distribution of

$$\frac{(1/d!)\sum_{t=1}^{n}\left(X_t - \sum_{i=0}^{d-1}(\nabla^i X_0)(t+i-1)^{(i)}/i! - \mu(t+d-1)^{(d)}/d!\right)}{\left[\gamma_Y(0)\sum_{v=1}^{n}\left[(v+d-1)^{(d)}\right]^2 + 2\sum_{h=1}^{n-1}\gamma_Y(h)\sum_{v=1}^{n-h}(v+d-1)^{(d)}(v+h+d-1)^{(d)}\right]^{1/2}} \tag{11}$$

converges to that of N(0,1).

Before developing a proof of Theorem 1, a few remarks and a corollary will help clarify this result. First, the centering and norming sequences in Theorem 1 are chosen to match the mean and variance of (11) with that of N(0,1) <u>for each $n \geq 1$</u>, not just in the limit. This tends to make the normal approximation more accurate. Because $\{Y_t\} \sim$ ARMA(p,q), and therefore Y_t has the representation (4), it can be shown that the series (8) converges absolutely, and

$$\gamma_Y(0) + 2\sum_{h=1}^{\infty}\gamma_Y(h) = \sigma^2\left(\sum_{j=0}^{\infty}\psi_j\right)^2,$$

where $\{\psi_j\}$ are the constants in the representation (4) with $\psi_j = 0$ for $j<0$ by causality. By elementary asymptotic analysis of sums of integer powers, both of the sums on v appearing in the denominator of (11) are asymptotic to $n^{2d+1}/(2d+1)$ as $n \to \infty$. Finally, the polynomial with stochastic coefficients that appears in the numerator of (11) leads, after summation, to a term that is equal to $\mu(n+d)^{(d+1)}/[(d+1)!]+O_p(n^d)$ as $n\to\infty$, again by simple asymptotic analysis of sums of integer powers. Hence, the centering constant in (11) can be modified, and the normalization constants simplified, yielding the following corollary.

Corollary 1. Under the conditions of Theorem 1, $n^{-1/2-d}\sum_{t=1}^{n}\left(X_t-\mu(t+d-1)^{(d)}/d!\right)\longrightarrow N(0,\omega^2)$ as $n\to\infty$, where

$$\omega^2=(2d+1)^{-1}d!^{-2}\left[\gamma_Y(0)+2\sum_{h=1}^{\infty}\gamma_Y(h)\right]. \tag{12}$$

The proof of Theorem 1 requires several lemmas. Ultimately, the goal is to express $\sum_{t=1}^{n}X_t$ as a weighted sum of the Y_t, which have the representation $Y_t=\sum_{j=0}^{\infty}\psi_jZ_{t-j}$, and then to exploit the fact that $\{Z_t\}\sim IID(0,\sigma^2)$ in order to apply a classical central limit theorem. The first lemma is fundamental in this goal.

Lemma 1. Suppose that $\nabla^dX_t=Y_t+\mu$, $t\geq1$. Then for $t\geq1$, X_t can be expressed as

$$X_t=\sum_{i=0}^{d-1}(\nabla^iX_0)\frac{(t+i-1)^{(i)}}{i!}+\frac{\mu(t+d-1)^{(d)}}{d!}+\sum_{v=1}^{t}Y_{t-v+1}\frac{(v+d-2)^{(d-1)}}{(d-1)!} \tag{13}$$

and for $n\geq1$

$$\sum_{t=1}^{n}X_t=\sum_{i=0}^{d-1}(\nabla^iX_0)\frac{(n+i)^{(i+1)}}{(i+1)!}+\frac{\mu(n+d)^{(d+1)}}{(d+1)!}+\sum_{v=1}^{n}Y_{n-v+1}\frac{(v+d-1)^{(d)}}{d!} \tag{14}$$

Proof: It may be assumed without loss of generality that $\mu=0$. The first formula follows from an induction argument on d. If d=1, then for $k\geq1$ $X_k-X_{k-1}=Y_k$, by definition, and summing this gives $X_t=X_0+\sum_{k=1}^{t}Y_k=X_0+\sum_{v=1}^{t}Y_{t-v+1}$, so (13) holds for d=1. Assume (13) holds for some $d\geq1$, and rewrite it as

$$X_t=\sum_{i=0}^{d-1}(\nabla^iX_0)(t+i-1)^{(i)}/i!+\sum_{v=1}^{t}\nabla^dX_{t-v+1}(v+d-2)^{(d-1)}/(d-1)!. \tag{15}$$

Suppose that $\nabla^{d+1}X_t=Y_t$. Then $\nabla^dX_t=\sum_{j=1}^{t}Y_j+\nabla^dX_0$. Substituting t-v+1 for t in this, and using (15), the induction hypothesis, it follows that

$$\begin{aligned}X_t &= \sum_{i=0}^{d-1}(\nabla^iX_0)(t+i-1)^{(i)}/i!+\sum_{v=1}^{t}\left(\sum_{j=1}^{t-v+1}Y_j+\nabla^dX_0\right)(v+d-2)^{(d-1)}/(d-1)!\\ &= \sum_{i=0}^{d}(\nabla^iX_0)(t+i-1)^{(i)}/i!+\sum_{v=1}^{t}\left(\sum_{j=1}^{t-v+1}Y_j\right)(v+d-2)^{(d-1)}/(d-1)!\\ &= \sum_{i=0}^{d}(\nabla^iX_0)(t+i-1)^{(i)}/i!+\sum_{v=1}^{t}Y_{t-v+1}(v+d-1)^{(d)}/d!,\end{aligned}$$

by applications of (9) and (10), a reversal of the order of summation, and a change of variable. This completes the induction proof of (13). Relation (14) follows by summing relation (13) from t=1 to n, reversing the order of summations, and using (9).

Lemma 2. Let $\{Z_t\} \sim \mathrm{IID}(0,\sigma^2)$, and for each $n \geq 1$, let $a_{t,n}$, $1 \leq t \leq n$, be a sequence of constants satisfying

$$\max_{1 \leq t \leq n} \frac{a_{t,n}^2}{\sum_{t=1}^n a_{t,n}^2} \rightarrow 0, \text{ as } n \rightarrow \infty. \tag{16}$$

Let $X_{t,n} = \dfrac{a_{t,n} Z_t}{\sigma\left[\sum_{t=1}^n a_{t,n}^2\right]^{1/2}}$, $1 \leq t \leq n$, and $S_n = \sum_{t=1}^n X_{t,n}$. Then as $n \rightarrow \infty$, $S_n \Rightarrow N(0,1)$.

Proof: $\sum_{t=1}^n E(X_{t,n}^2) = \sigma^2$ for all n. Let $\varepsilon > 0$. Then

$$\sum_{t=1}^n E\left(X_{t,n}^2 1_{\{|X_{t,n}| > \varepsilon\}}\right) \leq \sigma^{-2} E\left(Z_1^2 1_{\left\{Z_1^2 \max_{1 \leq t \leq n} a_{t,n}^2 \left[\sigma^2 \sum_{t=1}^n a_{t,n}^2\right]^{-1} > \varepsilon\right\}}\right) \rightarrow 0$$

as $n \rightarrow \infty$ by the dominated convergence theorem. Hence, by the Lindeberg - Feller Central Limit Theorem (Durrett (1991), p. 98), the result follows.

The final lemma needed in the proof of Theorem 1 is proposition 6.3.9 from Brockwell and Davis (1992). A sketch of its proof based on convergence of characteristic functions is given there. Here, a slightly different proof is presented in detail. For random k-vectors X_n and X, $X_n \Rightarrow X$ as $n \rightarrow \infty$ means that $P\{X_n \in A\} \rightarrow P\{X \in A\}$ as $n \rightarrow \infty$ for every k-dimensional Borel set A whose boundary ∂A satisfies $P\{X \in \partial A\} = 0$, which is equivalent to the condition that $Ef(X_n) \rightarrow Ef(X)$ as $n \rightarrow \infty$ for every bounded and continuous real valued function f on the k-dimensional real numbers. This is in turn equivalent to $Ef(X_n) \rightarrow Ef(X)$ as $n \rightarrow \infty$ for all bounded and uniformly continuous real valued functions f on the k-dimensional real numbers.

Lemma 3. Let X_n, n=1, 2, ... and Y_{nj}, j=1, 2, ...; n=1, 2, ... be random k-vectors such that

(i) $Y_{nj} \Rightarrow Y_j$ as $n \rightarrow \infty$ for each fixed j

(ii) $Y_j \Rightarrow Y$ as $j \rightarrow \infty$

(iii) $\lim_{j\to\infty}\limsup_{n\to\infty} P\{|X_n - Y_{nj}|>\varepsilon\}=0$ for every $\varepsilon>0$, where $|\bullet|$ signifies the usual Euclidean norm.

Then $X_n \Rightarrow Y$ as $n\to\infty$.

Proof: Let f be a bounded uniformly continuous real valued function defined on the k-dimensional real numbers. It is sufficient to show that $|Ef(X_n)-Ef(Y)|\to 0$ as $n\to\infty$. Fix an $\varepsilon>0$. There exists a $\delta>0$ such that $|f(x)-f(y)|\le\varepsilon$ for any x and y satisfying $|x-y|\le\delta$. By the triangle inequality,

$$|Ef(X_n)-Ef(Y)| \le E|f(X_n)-f(Y_{nj})| + |Ef(Y_{nj})-Ef(Y_j)| + |Ef(Y_j)-Ef(Y)|.$$

Denoting an upper bound of the function f by C, the first term on the right side of the inequality is bounded by

$$2C\, P\{|X_n - Y_{nj}|>\delta\} + \varepsilon.$$

It follows from (i) that for any j,

$$\limsup_{n\to\infty}|Ef(X_n)-Ef(Y)| \le \varepsilon + 2C \limsup_{n\to\infty} P\{|X_n - Y_{nj}|>\delta\} + |Ef(Y_j)-Ef(Y)|.$$

Taking the $\limsup_{j\to\infty}$ on both sides and using (ii) establishes the result, since $\varepsilon>0$ was arbitrary.

Proof of Theorem 1: By Lemma 1, the remarks following Theorem 1, and the fact that if $c_n\to c$ and $Z_n\Rightarrow Z$ as $n\to\infty$, then $c_nZ_n\Rightarrow cZ$ as $n\to\infty$, it suffices to show that

$$\frac{(2d+1)^{1/2} d! \sum_{v=1}^{n} Y_{n-v+1}(v+d-1)^{(d)}/d!}{\sigma\left|\sum_{j=0}^{\infty}\psi_j\right| n^{d+1/2}} \Rightarrow N(0,1) \qquad (17)$$

as $n\to\infty$. By representation (4) with $\psi_j=0$ for $j<0$ by causality, $Y_t=\sum_{j=0}^{\infty}\psi_j Z_{t-j}$ for all t. Define

$$Q^*_{nm} = \sum_{v=1}^{n}\left(\sum_{j=0}^{m}\psi_j Z_{n-v+1-j}\right)(v+d-1)^{(d)}/d!.$$

Changing indices of summation by letting $t=n-v+1$, and then $j=t-k$, it follows that

$$Q^*_{nm} = \sum_{k=1-m}^{-1}\left(\sum_{t=1}^{k+m}\psi_{t-k}(n-t+d)^{(d)}/d!\right)Z_k + \sum_{k=0}^{n-m}\left(\sum_{t=k}^{k+m}\psi_{t-k}(n-t+d)^{(d)}/d!\right)Z_k$$
$$+\sum_{k=n-m+1}^{n}\left(\sum_{t=k}^{n}\psi_{t-k}(n-t+d)^{(d)}/d!\right)Z_k.$$

Denote these last three double sums respectively by $Q^*_{nm}(1)$, $Q^*_{nm}(2)$, and $Q^*_{nm}(3)$. Then for fixed m, $Q^*_{nm}(1) = O_p(n^d)$, and thus $n^{-d-1/2}Q^*_{nm}(1) \xrightarrow{P} 0$ as $n\to\infty$. Also, for large n, $\mathrm{Var}(Q^*_{nm}(3)) \le \sigma^2 n^{2d}(\sum_{j=0}^{\infty}|\psi_j|/d!)^2$ so that $\mathrm{Var}\left(n^{-d-1/2}Q^*_{nm}(3)\right) \to 0$ as $n\to\infty$ and hence $n^{-d-1/2}Q^*_{nm}(3) \xrightarrow{P} 0$ as $n\to\infty$. Denoting the coefficient of Z_k in the second double summation by $a_{k,n}$, it follows by the summation-by-parts formula that

$$a_{k,n} = \left(\sum_{j=0}^{m}\psi_j\right)\frac{(n-k-m+d)^{(d)}}{d!} + O(n^{d-1}) \tag{18}$$

where the $O(n^{d-1})$ term is uniform in k, $0\le k\le n$, for fixed n. Thus, $\sum_{k=0}^{n}a^2_{k,n}=O(n^{2d+1})$, and hence (16) is satisfied, and by Lemma 2, and the fact that (18) implies that

$$\sum_{k=0}^{n}a^2_{k,n}=\frac{n^{2d+1}}{(2d+1)d!^2}\left(\sum_{j=0}^{m}\psi_j\right)^2 + O(n^{2d})$$

it follows that for each fixed m,

$$Q_{nm}=\frac{(2d+1)^{1/2}d!\,Q^*_{nm}}{n^{d+1/2}\sigma\left|\sum_{j=0}^{\infty}\psi_j\right|} \to N\left(0,\frac{\left(\sum_{j=0}^{m}\psi_j\right)^2}{\left(\sum_{j=0}^{\infty}\psi_j\right)^2}\right), \tag{19}$$

and obviously, as $m\to\infty$, the last random variable in (19) converges in distribution to N(0,1). To conclude (17) from Lemma 3, it is sufficient, by Chebychev's inequality, to show that

$$\lim_{m\to\infty}\limsup_{n\to\infty}\mathrm{Var}\left(Q_{nm}-\frac{(2d+1)^{1/2}d!\sum_{v=1}^{n}\left(\sum_{j=0}^{\infty}\psi_jZ_{n-v+1-j}\right)(v+d-1)^{(d)}/d!}{n^{d+1/2}\sigma\left|\sum_{j=0}^{\infty}\psi_j\right|}\right)=0.$$

Let $D(n,d) = \sigma|\sum_{j=0}^{\infty}\psi_j|\, n^{d+1/2}\,(2d+1)^{-1/2}d!^{-1}$. Notice that

$$E\left(Q_{nm}-\frac{(2d+1)^{1/2}d!\sum_{v=1}^{n}\left(\sum_{j=0}^{\infty}\psi_jZ_{n-v+1-j}\right)(v+d-1)^{(d)}/d!}{n^{d+1/2}\sigma\left|\sum_{j=0}^{\infty}\psi_j\right|}\right)^2 =$$

$$E\left(\frac{\sum_{v=1}^{n}\frac{(v+d-1)^{(d)}}{d!}\sum_{j=m+1}^{\infty}\psi_jZ_{n-v+1-j}}{D(n,d)}\right)^2 =$$

$$D^{-2}(n,d)\sigma^2\Bigg[\sum_{j=m+1}^{\infty}\sum_{k=m+1}^{\infty}\psi_j\psi_k\sum_{v=1}^{n}\left(\frac{(v+d-1)^{(d)}}{d!}\right)^2 +$$

$$2\sum\sum_{m+1\le k<j}\psi_j\psi_k\sum_{v=1+j-k}^{n}\frac{(v+d-1)^{(d)}(v+d-j+k-1)^{(d)}}{d!^2}\Bigg] \to \frac{\left(\sum_{j=m+1}^{\infty}\psi_j\right)^2}{\left(\sum_{j=0}^{\infty}\psi_j\right)^2},$$

as $n\to\infty$. The result now follows by letting $m\to\infty$. The final form (11) of Theorem 1 follows by using (13) and (14) to verify that (11) has mean 0 and variance 1.

Conclusions

A central limit theorem has been developed for centered and normalized sums of random variables that constitute an ARIMA(p,d,q,) process. If $\{X_t\}$ is an ARIMA(p,d,q) process satisfying $\nabla^d X_t=\mu+Y_t$ with $\{Y_t\}\sim$ARMA(p,q), it is seen that the proper normalization and centering sequences in $b_n^{-1}\sum_{t=1}^{n}\left(X_t-a_t\right)$ are $b_n=n^{d+1/2}$ and $a_t=\mu(t+d-1)^{(d)}/d!$.

Among other applications, this central limit theorem is important in making large sample approximations related to sums of the form $\sum_{t=1}^{N_T} X_t$ where $\{N_T, T\geq 0\}$ is a nonhomogeneous Poisson process and $\{X_t\}$ is an ARIMA(p,d,q) process, independent of $\{N_T, T\geq 0\}$. This model provides a realistic representation of the total claims cost associated with medical claims / compensation costs associated with accidents or hazardous materials exposure.

References

Box, G.E.P. and Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control. Holden Day: San Francisco.

Brockwell, P.J. and Davis, R.A. (1992). Time Series: Theory and Methods. New York: Springer-Verlag.

Durrett, R. (1991). Probability: Theory and Examples. Pacific Grove: Wadsworth & Brooks/Cole.

Prabhu, N.U (1980). Stochastic Storage Processes: Queues, Insurance Risk, and Dams. New York: Springer Verlag.

Priestley, M.B. (1981). Spectral Analysis and Time Series, Vols. I and II. New York: Academic Press.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
August 1992

3. REPORT TYPE AND DATE COVERED
Final June 92 - August 92

4. TITLE AND SUBTITLE
A Central Limit Theorem for Autoregressive Integrated Moving Average Processes

5. FUNDING NUMBERS
Program Element:
Work Unit Number:
NAVFAC Reimbursable

6. AUTHOR(S)
Dr. John E. Angus

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92186-5122

8. PERFORMING ORGANIZATION
Report No. 92-15

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Medical Research and Development Command
National Naval Medical Center
Building 1, Tower 2
Bethesda, MD 20889

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
This research was undertaken on an ASEE Summer Faculty Research Fellowship

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

A modification of the classical compound Poisson process has been shown to be a viable model for describing total claims and compensation costs associated with hazardous material exposure. This model assumes that the successive claim awards form an autoregressive integrated moving average (ARIMA) process. In order to make practical approximations to the distribution of total claim and compensation costs with this model, it is necessary to develop approximations to the distribution of the sum of successive elements from an ARIMA process. In this investigation, a central limit theorem is developed for ARIMA processes, that makes this approximation possible in the case where the number of summands is large.

14. SUBJECT TERMS
ARIMA processes, compound Poisson process, nonhomogeneous Poisson process, claims and compensation costs, normal approximation, time series

15. NUMBER OF PAGES
13

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited